Efficient/Pareto Frontier
(Notes on a Pair of R functions)

When two or more dimensions of quality, typically cost and benefit, have to be considered and cannot be traded off one against the other by a simple tariff or weighting, the concept of the *Efficient Frontier* allows us to identify options that remain feasible whatever weight we place on one dimension of value or another. It is also often called the *Pareto Frontier*, in honour of economist Vilfredo Pareto (1848-1923) although it doesn't seem to have appeared explicitly in his work.

The basic idea is further described in the following Wikipedia article on multi-objective optimization.
https://en.wikipedia.org/wiki/Multi-objective_optimization

Beware of Youtube videos describing the Efficient Frontier with reference to portfolio management. The ones I have seen misrepresent the core concept!

I have written a couple of little R functions that allow me to identify (& plot) the "efficient" or Pareto frontier in a group of items where one wants to maximize or minimize two dimensions but doesn't have a tariff for comparing quality on one dimension to the other. Typically this is used with a cost and a benefit score to identify feasible solutions, out of which one may be selected on the basis of other criteria. (These R functions only work with 2 dimensions of value, which is normal. Using more dimensions complicates the graphics and usually leaves too many items on the frontier, so with inherently high-dimensional problems some preparatory weeding out or combining of value-dimensions is usually done in practice to arrive at a 2-dimensional problem.)

The function frontier()
```
frontier = function(xvec,yvec,namevec=",xupward=T,yupward=T)
```
takes two value vectors, xvec and yvec, as input, along with an optional namevec of labels and identifies those items that lie on the efficient frontier, returned as a data frame. The two parameters xupward and yupward control whether maximization or minimization should be applied to the x and y values separately. Hence there are four possible modes.

The function frontout()
```
frontout = function(datframe,xcol,ycol,namecol=",xupward=T,yupward=T)
```
takes a dataframe and two column names and uses frontier() to plot a graph of the options under scrutiny and highlight those on the efficient frontier in a graph.

An example plot is shown on the next page.

This example isn't typical of how efficient frontiers tend to be used in economics or decision-making, but it illustrates the general idea quite well. It seeks world-war-2 aircraft that are both fast and have long range, i.e. with 2 factors to be maximized rather than a cost-benefit trade-off. Items on the frontier are named. The unnamed black dots are 'dominated', meaning that one or more item on the frontier is both faster and has longer range. (No British planes achieved a place this frontier, not even the much-loved Supermarine Spitfire.)

WWII aircraft: maxrange & topspeed 2b maximized.
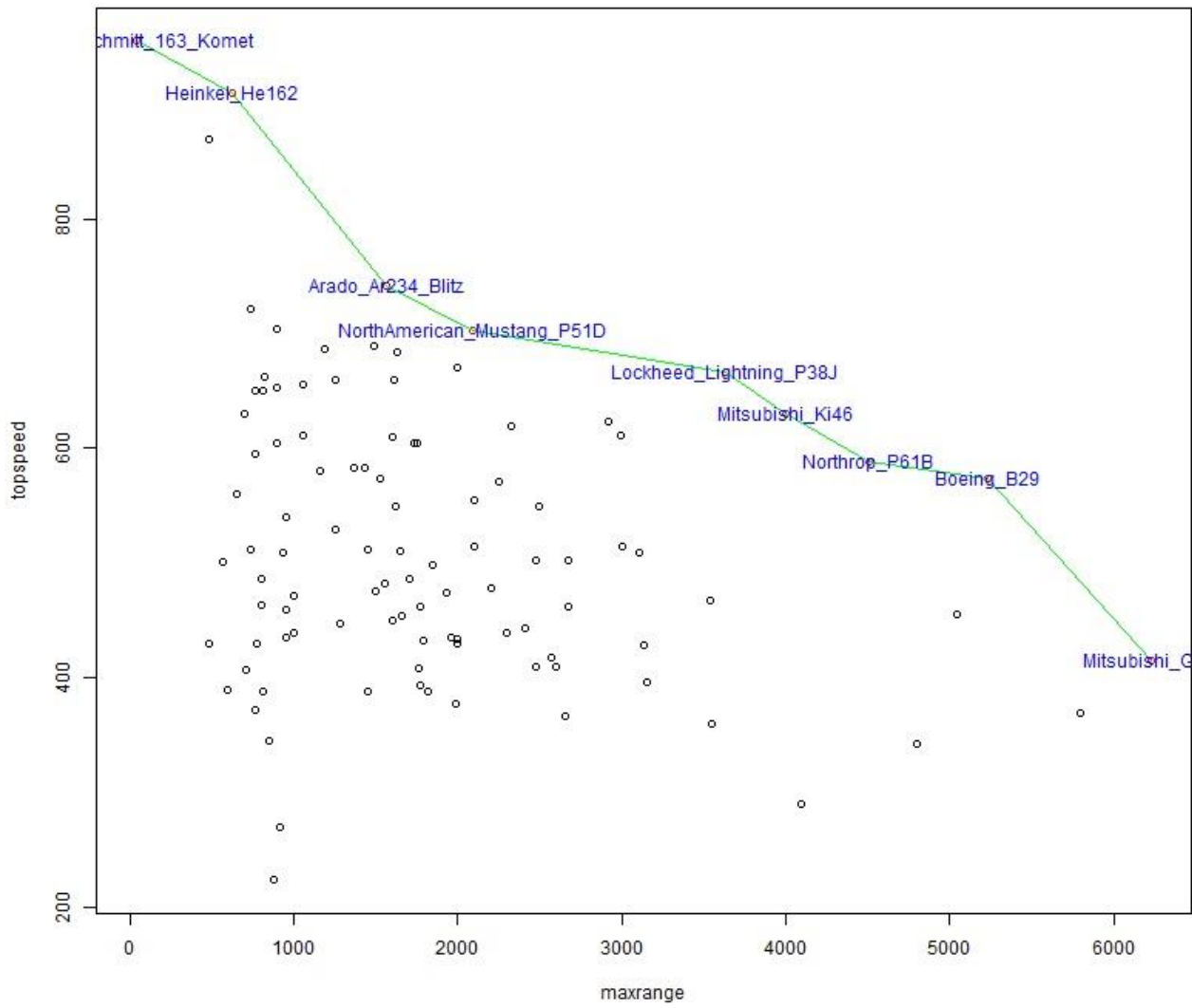
[Richard Forsyth, September 2020.]